

DOCUMENT RESUME

ED 122 776

IR 003 390

AUTHOR Tennyson, Robert; And Others
TITLE Evaluation: Its Role in Instructional Development.
PUB DATE Apr 76
NOTE 21p.; Paper presented at the Annual Meeting of the Association for Educational Communications and Technology (Anaheim, California, March 28-April 2, 1976)

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage
DESCRIPTORS *Evaluation Methods; Formative Evaluation; Guidelines; *Instructional Design; *Instructional Materials; *Material Development; *Models; Summative Evaluation

IDENTIFIERS AECT 76

ABSTRACT

An instructional development model is described that provides for evaluation of the feasibility of a project, evaluation during the formative design and development phases, summative evaluation prior to dissemination, and continued evaluation after implementation. These guidelines view evaluation as an integral part of development; before the output of one phase goes into the next it is thoroughly evaluated. (JY)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

Presented at the AECT Convention, 1976, California

Evaluation: Its Role in Instructional Development

Robert Tennyson, Carol Tennyson, and Daniel Dobbert

University of Minnesota

Conventional instructional development models are composed of sequential steps which lead to the production of instructional materials (cf. Dick, 1974). The steps of instructional development most often include these phases (see Figure 1): identification of goals and objectives, design of instruction materials, development and production, and evaluation (Briggs, 1970). The rationale often used for placing evaluation last in the development process is that it provides a method for revising the instructional product prior to dissemination. However, evaluation of the materials at that point can be too late for cost-effective product revisions because the repetitive recycling loop (see Figure 1) comes when the product is basically ready for dissemination (Merrill, 1973). Thus there is a need for an instructional development model that provides for evaluation of the feasibility of the project, evaluation during the formative design and development phases, summative evaluation prior to dissemination and provides continued evaluation after implementation. The purpose of this paper is to present instructional materials evaluation guidelines that eliminate the assumption of a recycling evaluation system in instructional development.

Insert Figure 1 about here

Four Phases of Evaluation

Evaluation of instructional materials involves four basic interactive phases (Figure 2). In the first phase, feasibility, the appropriateness of

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

the procedures used to analyze the instructional need and development capabilities is evaluated, while in the second formative, the instructional materials are evaluated at various points during the design and development phases for purposes of revisions and refinements. When the instructional materials have been developed, evaluation in the summative phase determines the extent to which the materials adequately enable the learner to effectively and efficiently perform the behaviors being taught. In the final phase, maintenance, the instructional materials are evaluated for purposes of adoption and/or modification, and after implementation, evaluates the instructional materials for possible continuing use.

Insert Figure 2 about here

Feasibility Evaluation

The first evaluation activity occurs concurrently with the instructional development problem analysis phase (see Figure 2). In this developmental phase teachers or other potential developers should first identify the instructional problem and then provide recommendations concerning the need for developing instructional materials. If the decision is made to develop the instruction then the design and development specifications can be established in subsequent phases. The purpose of the feasibility evaluation is to evaluate the procedures and rationale used in the analysis phase (Scriven, 1975).

Guidelines for conducting a feasibility evaluation are listed below.

1. Problem identification methods are fairly standard and the evaluation here focuses on documenting these procedures. The source of the data used in the needs assessment should be specified. In other words, how and from whom was the data collected? For example, if data are from

school faculty (or other experts) were they polled quantitatively or asked opinions, if from societal sources, were individual representatives or organizations contacted; if from educational policies were the policies consistent with known research findings or hoped for results?

2. Goals and objectives should be directly related to the specified needs as identified in the problem analysis. Document the source of both the goals and objectives.
3. Resources and other situational variables which control the specifications of the development effort can be evaluated in terms of cost-benefit. Will the expenditure return a product that meets the goals and objectives of the project? Document completely the estimated costs and sources of funding (if needed) and the availability of development and production capability.

Formative Evaluation

The purpose of a formative evaluation is to obtain data necessary for making revisions and refinements of the instructional materials during the design and development phases (see Figure 2). Data collected for this evaluation are derived from qualitative and quantitative sources. These sources are mainly subject matter consultants and learners for whom the instruction is intended.

Content review. Part one of the formative evaluation starts during the design phase with a review and critique of the content structure by subject matter consultants. The content structure refers to the scope and sequence of the subject matter to be included in the instructional materials. This procedure is more qualitative than quantitative. Much of the content that is learned is determined by consensus. Having subject matter consultants review

the content and offer suggestions can help assure that the content structure reflects that consensus. This form of evaluation is particularly useful in making refinements on the content since learners, because of their naiveté, often contribute little to this part of the evaluation.

A review by subject matter consultants is challenging because the teacher (developer) has to accept criticism of the instructional materials and be willing to work to make any appropriate revisions and refinements. Persons selected for the review should give as complete interpretations of the materials as possible. Reviewers need assurance that their opinions will be received as comments for improvement, not as responses to be evaluated as good or bad. If the latter happens, the evaluation data are definitely not acceptable for the purposes of revision and refinement.

In designing instructional materials it is necessary to determine the sequence in which the content is to be presented. The sequence is in part controlled by the nature of the content and by the behaviors required of the learner in relationship to that content. Sequences fall into three basic types—hierarchical, algorithmic, and random (Tennyson, 1972). In hierarchical sequences, lower order behaviors need to be learned before efficient learning of higher order (for example, to be a good problem solver the learner needs to learn concepts and rules which can be applied to solve problems). In algorithmic (information processing) sequences, the output of one step becomes the input of the next step (for example, statistical formulas follow a sequence such that if a part were left out it is not possible to apply the formula correctly). In random sequences, the content and behaviors can be learned in any order although certain "logical" orders might be more efficient.

Guidelines for evaluating the content and behavioral sequence include:

1. Subjectively determine which of the three types of sequences the content and behaviors fall into.
2. For hierarchical sequence, develop a test which will assess each level of the hierarchy. Administer the test to at least two different groups of people: one group that is assumed to have some of the behaviors and a second assumed to have all of the behaviors. Test results should then be graded according to levels in the hierarchy. Items identified correctly most frequently should correspond to the bottom of the hierarchy while items most frequently missed should be at the top. Where necessary the hierarchical sequence can be adjusted with these results (Tennyson & Boutwell, 1971).
3. To check an algorithmic sequence, design two tests. In both cases give the algorithm or steps for reference, but for one test either leave out a step or put the steps in a different order, and for the second give the correct sequence. Learners receiving the "incorrect order" or "missing step" test should not be able to perform as readily as the group receiving the complete algorithm.
4. For a random sequence, the same procedures followed in the above two sequences can be used since if the test results show neither a hierarchy nor an algorithm than a random sequence can be assumed. Therefore, the structure of the content should be thoroughly studied for determining an efficient, logical order.

One-to-one tryout. The second part of the formative evaluation takes place concurrently with the development of the instructional materials (see Figure 2). This procedure involves a tryout of the product between the teacher (developer) and a person for whom the instruction is intended. While

the learner works through the instructional materials the teacher observes and takes notes. The purposes of the one-to-one tryout are to identify editorial problems, to identify any unclear directions, and to identify those contents areas which are unclear.

When conducting a one-to-one evaluation the following guidelines may be helpful.

1. Use mimeo, ditto, or photo-copies, and encourage the learner to make written comments directly on the material.
2. Make sure the learner is aware of his/her role in the tryout. Encourage the learner to be critical--not just complimentary.
3. Where appropriate work with the learner to clarify segments of the instruction which are not clear.
4. Keep extensive notes on learner comments as well as your own observations.
5. Ask the learner to indicate confidence in his/her answers on practice portions of the instruction and tests. For example, one method of confidence rating is the following three choice response. The student is asked to mark the appropriate response.

- | | |
|----------------|---|
| A. very sure | if you think that your chances of being right are better than 50:50 |
| B. fairly sure | if you think that your chances of being right are just about 50:50 (for example, if you can't decide between two answers) |
| C. guess | if you think that your chances of being right are less than 50:50 |

On each test question the learner rates his/her response confidence. This data is useful in estimating retention of the content (Alhgren, 1969). If a learner has a good score and a high confidence rating then it is likely that he/she will remember the content; a good score and low confidence is an indication that the learner understood the

content for the test but will likely not retain the content over time. A good score and a low confidence rating is also an indication that the instruction was inadequate. A low score and high confidence rating would indicate a faulty test key, faulty instruction, or poor learner motivation or attitude.

6. When you have completed the one-to-one tryout, study the data for possible revisions of the instruction. If numerous changes are required conduct a second one-to-one tryout after the revisions have been made. All major revisions of the instructional materials should be completed before the final part of the formative evaluation process begins.

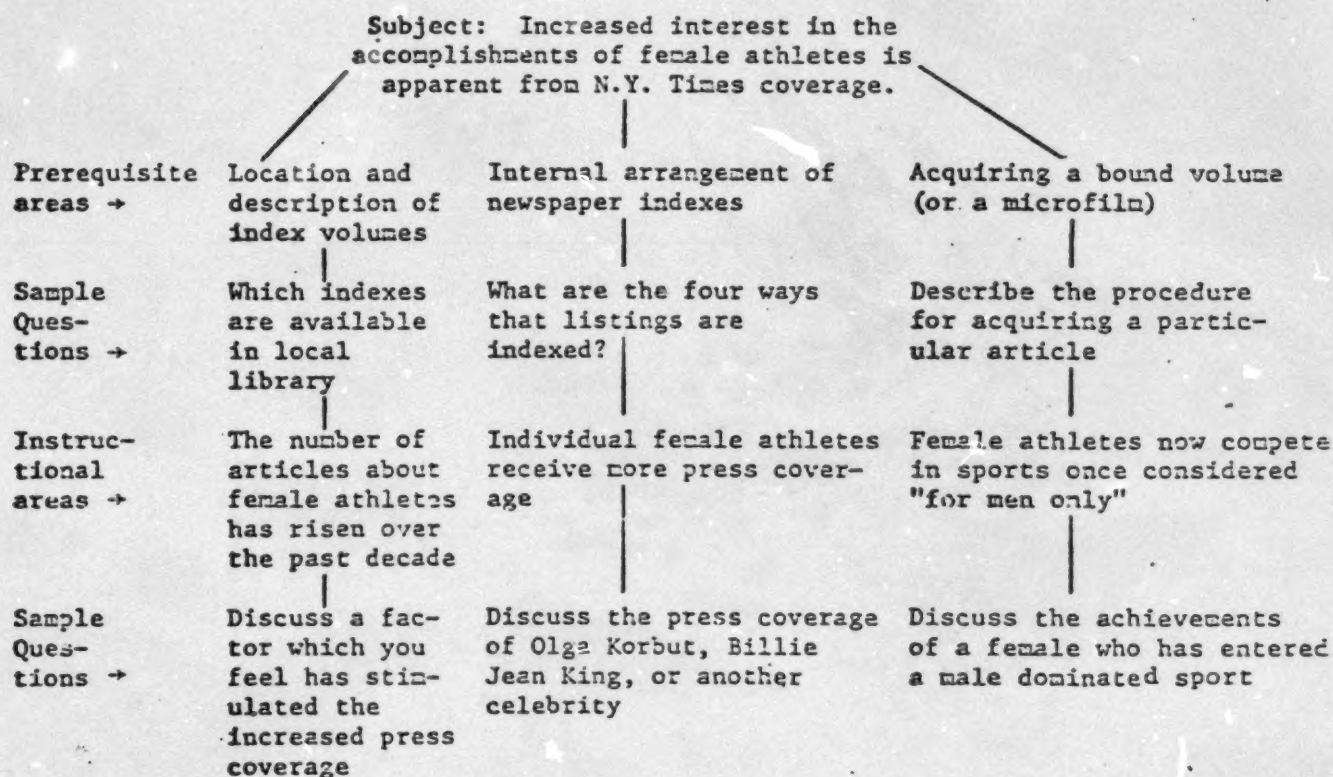
Group tryout. The final step in formative evaluation involves a group tryout (see Figure 2). The purpose is to obtain information for making any major refinements before beginning production of the materials. Variables to be considered for refinement here include learner time for direct use of the instructional materials and indirect study (e.g., homework), pretest scores to assess prerequisite skill levels, and pre test/posttest scoring patterns.

A group can include from 5-10 learners. Refinements are usually changes (either deletions or additions) resulting from variations in the learner population. Certain learners may need additional instruction or perhaps more help in understanding certain parts of the content, or it could be that some amount of practice is redundant and should be left out. In all, refinements in the instructional materials represent changes associated with individual differences found in the learner population. Certainly, the identification of such differences that might necessitate refinements is only possible with a group tryout.

The following numbered paragraphs may be considered as guidelines for the group evaluation.

1. If possible learners should be assured of anonymity and encouraged to make written comments. Confidence ratings are also recommended. In situations where the group tryout will be graded, it should be recognized that certain personal data, such as attitudes toward the instruction and confidence ratings, can be collected only if learner anonymity is assured.
2. Scheduling and resource availability, although not covered in this paper, are directly connected to the success of the instruction, especially with regard to learner attitudes. Management of the learning environment requires attention to details to make sure all resources and facilities are readily accessible to the learners.
3. Data should be collected on the amount of time each learner spends using the instructional materials directly and the amount of time each learner spends with outside resources. All learners should be directed on how to record time spent with various resources and assured that this information will not affect their grades in any way. Time data makes it possible to assess the efficiency of the instructional materials. For example, if learners spend a lot of time on the instructional materials but have only marginal performance then something is wrong with the instruction. The best situation is minimal time and maximum performance. Time recording in minutes is sufficient for most instructional programs.
4. Learner performances on the pretest and posttest are useful in determining adequacy of the instruction. The pretest has the dual function of assessing prerequisite skills and measuring prior

knowledge of the content to be learned in the instruction. Except for items measuring prerequisite skills, the pretest and posttest should be equivalent. That is, the two tests should be defined as A and B, and the sequence as to which is pre and post can be either A and B or B and A. The following example provides an illustration of a pretest which includes testing of both the prerequisite and instructional behaviors. The content areas and sample test questions are from an introductory course in journalism.



Learning from instructional materials can be inferred if students show a significant gain in performance from the pretest to the posttest. The simplest method of determining gain, is to look at the difference between the pretest and posttest scores. Statistical procedures such as the paired t-test or Walsh test are usually not used here because the sample size is small and individual

scores are still considered more important than group means. Test questions are usually developed in reference to the objectives and/or content of the instruction; that is, each pretest and posttest item can be directly related to specific instructional units (objectives). The instructional materials can be evaluated for effectiveness by comparing each learners pretest and posttest performance scores for each unit of instruction. There are four possible response patterns for diagnosing potential problems in the instructional materials, and these are:

<u>Pretest</u>	<u>Posttest</u>
item missed	item passed
item missed	item missed
item passed	item missed
item passed	item passed

In this illustration the test items for the pretest and posttest are equivalent—not necessarily the same item. The first response pattern of missed/passed indicates learning; probably no refinements of the instruction are required. But, for the second and third patterns, no learning can be inferred. And, of these two, the passed/missed pattern, could be the more serious problem. To diagnose the instruction for revisions the scoring key should be checked first. If it is correct, then a careful analysis of the instruction must be done. Most likely, a misconception of that part of the content has occurred. Pattern four seems to imply that the learner knows the content and behavior of that particular instructional unit. Therefore, either that segment may be dropped, or a provision to allow learners to by-pass instruction passed on the pretest may be included.

The data collected from the review and two tryouts discussed above should provide adequate information for revising and refining the instructional

materials. The instructional materials can be considered complete at this stage. The final step in evaluation should be done prior to dissemination to determine if the instructional materials enable students to meet the goals and objectives of the instruction.

Summative Evaluation

Third evaluation phase, summative, takes place in the environment or setting where the instruction is to be normally used. Its purpose is to determine to what degree the instructional materials meet the goals and objectives of the development project.

The basic instructional goal to be evaluated states that the learners will learn the defined content and behaviors in an effective and efficient manner. Data for this evaluation should include student performance, gain in learning, student time expended in learning, learner attitudes, comparison with a control group, and costs.

Guidelines for the summative evaluation may include the following:

1. The main consideration in conducting a summative evaluation is to tryout the instructional materials in the learners conventional learning environment. Learners should be told that the instructional materials are part of their usual learning experiences. That is, no unusual preparation or announcements are necessary. This is very important because learner reactions to experimental or different learning situations can influence their normal behavior.
2. Again, time data collected on direct use of the instructional materials and other learning sources should be collected if possible. A portion of the instructional goals and objectives focuses on efficiency of the instruction for learning. Time data should include information such as: (a) time used for directions, (b) time spent on

the different learning parts of the instructional materials, (c) time needed for outside study, and (d) total time. Learners can keep good records of their study time if proper directions are given and if the time logs are frequently checked and used by the teacher (developer).

3. Attitudes of the learners toward the instruction should be of special importance when the instructional materials introduce new methods of presentation (e.g., computer-based instruction) and/or deviates from conventional instruction. The most convenient method of gathering learner attitudes is by administration of a survey instrument. However, other methods of measuring attitudes should be considered also. Attitudinal responses should be specific to definable characteristics or variables of the instruction. Overall composite ratings are rarely useful in assessing quality of the instruction except in situations where there is clearly a negative trend. In most cases, attitudinal data should be collected from instruments and methods designed specifically to evaluate the instruction under development.
4. The most important question at this point is, "do the instructional materials teach?" There is only one way to really test this question, and that is to compare learners who have taken the instruction with those who haven't. There are variations of this control group design and many sources which can illustrate them. The point here is that some form of multiple group comparison is necessary to complete a summative evaluation. The usual design format is:

Group 1	Pretest	Instructional Materials	Posttest
Group 2	Pretest	No Instructional Materials	Posttest

The control group receives the pretest and posttest but not the instruction. The instructional materials group receives both the

tests and the instruction. At this point group means (average scores of learners) are studied rather than individual scores as in the formative evaluation. Also, the two groups should be samples from the same population. This can be checked by seeing if the pretest mean scores are similar. There should be no basic differences between the two groups on the pretest. However, on the posttest the mean scores should be significantly different. If the group sizes are large enough a statistical test can and perhaps should be done.

As mentioned above, different parts of the instructional design can be tested with this multiple group comparison design. Simply leave out of the instructional materials items that are questionable and test the remaining design variables against the complete materials. The design is:

Group 1	Pretest	Full Instruction	Posttest
Group 2	Pretest	Partial Instruction	Posttest

Of course, it is possible to include additional partial groups; only the number of questionable design variables and number of learners limit the number of groups.

In summary, at least the first comparison should be done. The others may be conducted if certain design alternative are being considered.

5. A final consideration in summative evaluation is a determination of costs involved in the development. Costs in many cases are "best-guess" estimates because personnel expenses are variable, overhead expenses are absorbed by other units and budgets (and, therefore, are a fixed cost regardless of how much was actually used), and media resource expenses for development and production are susceptible to

local policy controls and many other factors that can either inflate or understate the actual costs. The reason for this less than accurate method of calculating costs is that most development activities are supported by agencies or facilities outside the control of the teacher or developer. This is true whether it is in educational or industrial institutions; therefore, it is important to estimate costs as accurately as possible and to provide good documentation of how costs were determined. When the costs are finally calculated, the anticipated life of the product should be derived. This process seems to be a subjective analysis but the usual life expectancy of instructional materials without major changes is five years. By dividing the total costs by the years of life expectancy it is possible to amortize costs and absorb the costs yearly rather than all at once. By adding to each yearly development costs the operational costs of that year and dividing that by the number of learners for that year, the cost per student can be calculated. Other factors can be introduced into this method of determining the cost-effectiveness of the instruction. Obviously student learning might be included if it is possible to compare gain in effectiveness or efficiency with previous programs or in relationship to the goals and objectives.

Following the summative evaluation the instructional materials can be put into standard use, either in the classroom by the teacher or through commercial production by the developer. Another step in the instructional development model is the dissemination of the materials. This means that someone besides the developer or teacher is given the opportunity to use the materials. It is at this point that the fifth phase of evaluating instructional materials should begin.

Maintenance Evaluation

Evaluation of instructional materials after implementation is the third phase of evaluation. This phase is usually not considered in the conventional instructional development model, but it can influence the actual useful life of the materials. The purpose here is to maintain the instructional materials at or near the level of effectiveness as when first implemented. Guidelines to be considered for maintenance evaluation are presented in the following numbered paragraphs:

1. Are the instructional materials still worth using in the learning environment? This question refers to the concept of cost-benefits. In other words, do the benefits derived from the product justify the costs? In the case of instructional materials, are they achieving the goals and objectives of the learning environment? Benefits include such factors as high learning levels, positive learner attitudes, and an efficient management system. Probably the most important source of data to be analyzed here would be the performances of learners after they have finished the instruction. This can be immediate performance, such as performance at the next level of the learning environment, or performance after transferring into a non-instructional environment. Collection of this type of data is useful in updating the goals and objectives of not only the instructional materials but the entire learning environment.
2. Updating the content is an important consideration in keeping instructional materials current. Obviously some disciplines are more susceptible to changes than others. The teacher or developer should set up a procedure to review the content in light of changes in the field and to update the content accordingly. The assumption that

content in any field will remain constant for five years (or whatever) without updating is incorrect. It should be assumed that periodic changes need to be made.

3. Learner attitudes toward the instructional materials should be measured along with performance measures. Learner attitudes do fluctuate; they need to be reassessed after instructional materials have been used over a period of time. Factors which usually affect student attitudes are out-of-date visuals, poor condition of photographic displays (e.g. film and film strips), missing learning materials, or any other components of the system which do not meet the technical standards of the original product. Student attitudes of this kind are obtained not by questionnaires, but by one-to-one interviews and small group discussions; where possible anonymous inputs solicited by the teacher are helpful.
4. Changes in individual characteristics of the learner need to be evaluated. Certainly schools, and even industry, have had to reassess the goals of basic educational needs of students. Societal policies change periodically requiring that instructional goals and objectives change accordingly. Likewise, learner goals fluctuate, necessitating adjustments in learning environments and possibly changes in the instructional materials. Another form of learner imposed changes comes from shifts in learner characteristics. For a variety of reasons learner prerequisites change over time requiring possible changes in the instructional system.
5. If some form of technology is involved in the instructional delivery system, that must be evaluated and maintained as well. New media sources should be incorporated when possible. For example,

computerizing tests could provide additional years of usage to instructional materials. Integrating media into existing systems could improve the efficiency of the learning. Modification of the delivery system is often an inexpensive method of updating instructional materials.

Maintenance evaluation makes it possible to prolong the life of developed instructional materials while maintaining their original effectiveness and efficiency. To make significant gains in developing additional instructional materials requires that a method of maintaining current materials be used.

Summary

Evaluation of instructional materials begins with the very first phases of development and continues until they are finally replaced. The purpose of continuous evaluation for instructional development is not only to insure effective instruction but also to keep development costs and time minimal. Conventional instructional development models have placed evaluation at the end of the development process, requiring a recycling for revisions. Certainly this form of raw empiricism is costly and time consuming. The guidelines presented here view evaluation as an integral part of development. Before the output of one phase goes into the next it is thoroughly evaluated; recycling is not required. Obviously, the goal of the procedure is to save time and increase the likelihood that additional instructional materials will be developed.

References

- Ahlgren, A. Reliability, predictive validity, and personality bias of confidence-weighted scores. Confidence on achievement tests -- theory and applications. In R. Ebel (Chair), joint symposium presented at the meeting of the American Educational Research Association and National Council on Measurement in Education, Los Angeles, 1969.
- Briggs, L. J. Handbook of procedures for design of instruction. Pittsburg, Pa.: American Institutes for Research, 1970.
- Dick, W., An overview of the systems approach. In R. N. Singer and W. Dick. Teaching physical education, a systems approach. Boston: Houghton-Mifflin, Inc., 1974.
- Merrill, M. D. Content and instructional analysis for cognitive transfer tasks. AV Communications Review, 1973, 21, 109-125.
- Scriven, M. Evaluation prospects and procedures. In W.J. Popham (Ed.) Evaluation in education: Current applications. Berkeley, Calif.: McCutchin, 1975.
- Tennyson, R. D. A review of experimental methodology in instructional task sequencing. AV Communications Review, 1972, 20, (2), 147-159.
- Tennyson, R. D., & Boutwell, R. C. A quality control design for validating hierarchial sequencing of programmed instruction, NSPI Journal, 1971, 10, 5-10.

